

Cooperative Equilibrium beyond Social Dilemmas: Pareto Solvable Games

Valerio Capraro* Maria Polukarov† Matteo Venanzi† Nicholas R. Jennings†

September 28, 2015

Abstract

A recently introduced concept of *cooperative equilibrium* [7], based on the assumption that players have a natural attitude to cooperation, has been proven a powerful tool in predicting human behaviour in social dilemmas. In this paper, we extend this idea to more general game models, termed *Pareto solvable* games, which in particular include the Nash Bargaining Problem and the Ultimatum Game. We show that games in this class possess a unique pure cooperative equilibrium. Furthermore, for the Ultimatum Game, this notion appears to be strongly correlated with a suitably defined variant of the Dictator Game. We support this observation with the results of a behavioural experiment conducted using Amazon Mechanical Turk, which demonstrates that our approach allows for making statistically precise predictions of average behaviour in such settings.

1 Introduction

Over recent decades, numerous experimental studies of human behaviour in economic scenarios have demonstrated the implausibility of traditional game-theoretic solutions based on payoff maximisation, rational choice and equilibria [6]. One of the most influential examples in experimental economics, offering evidence against such rationally self-interested play, is the Ultimatum Game [23].¹ In this model, one player proposes a division of a given budget, b , between himself and the other player, who can either accept or reject the proposal. If the recipient accepts, the budget is split accordingly; if he rejects, both players get nothing.

Clearly, if the recipient is a payoff maximiser, he would accept any non-zero offer, so the proposer should offer the smallest amount possible. However, as experimental results show (see, e.g., [23, 29, 6, 31, 43, 46, 41, 44]), human subjects tend to deviate from this equilibrium behaviour. Specifically, the recipients often reject low offers, and the proposers often propose higher shares. While there is a considerable quantitative variation across studies, the following regularities can be observed:

- I. There are virtually no offers above $0.5b$, where b denotes the given budget [19].
- II. The vast majority of the offers lie in the interval $[0.3b, 0.5b]$ [40].
- III. There are almost no offers below $0.2b$ [19, 40].
- IV. Offers below $0.25b$ are rejected with high probability [18].

These results inspired several attempts to develop suitable models to explain human behaviour more accurately. Some were based on strategic arguments along the idea that higher offers are more likely to be accepted [41, 44]; others used the so-called “inequity aversion” model, postulating that people are averse to unfair situations and so propose a fairer budget split [19, 25, 32]. In this line, the classical Fehr-Schmidt inequity aversion model [19] is relatively successful in explaining the regularities I–IV. However, the parameter estimation method in this model (involving two parameters for each of the

*Centrum Wiskunde & Informatica (CWI), The Netherlands

†University of Southampton, United Kingdom

¹Other examples include the Prisoner’s Dilemma, the Public Goods Game, and the Dictator Game (see, e.g., [6])

players—see Appendix for more details) has come under serious criticism in the recent literature [2] due to the difficulty of generalising this method to a wider set of economic games. Furthermore, the recently proposed concept of “benevolence” (that is, making decisions towards increasing the opponent’s payoff beyond one’s own) [11], has challenged the very idea of inequity aversion (see also [14, 16]). Alternatively, Rand et al. [40] used stochastic evolutionary game theory to show that natural selection favours fairness. However, while their model showed excellent results in matching population averages, there was a wider variance in individual strategies than is normally present in empirical data. In particular, their model yields a large number of individuals who make and demand offers above 50 per cent of the budget [20].

Based on these findings, Camerer [6] suggested that human players in the Ultimatum Game are most likely to be motivated by a combination of strategic incentives and inequity aversion. In this work, we explore this idea and propose a solution concept that compounds these two factors. We note that, even though Fehr and Schmidt [19] also based their model on a similar approach, our work differs from theirs in a number of conceptual aspects, which in particular include the way we define the utility function. Specifically, in our model the strategic component comes in the form of a tendency to maximise a utility depending on the social welfare, while the inequity aversion component corresponds to a donation of the player in a certain variant of the Dictator Game [29] that we associate with the given instance of the Ultimatum Game. Both these behavioural tendencies have been reported by previous experimental studies [27, 22, 33, 39, 10, 11, 18, 5, 3, 4, 13, 15, 21].

In more detail, our model builds on the concept of cooperative equilibrium, previously proposed for two-player one-shot social dilemmas [7], iterated two-player social dilemmas [12], and n -player one-shot highly symmetric social dilemmas [1, 9]. We extend the definition to more general game classes, which we term here Pareto solvable games, due to their properties. These games in particular include the aforementioned Ultimatum Game, as well as other fundamental models, such as the Nash Bargaining Problem and the Coordination Game. We show that these games possess a unique cooperative equilibrium.

Importantly, cooperative equilibrium in the Ultimatum Game gives a solution that explicitly predicts the regularities III and IV observed in previous experiments. Moreover, this solution determines a precise correlation between a player’s behaviour in the Ultimatum Game and its corresponding variant of the Dictator Game. Using this correlation, we can give an a posteriori explanation for regularities I and II.

Essentially, the idea behind the concept of cooperative equilibrium is to reduce the original game to another game, called the *induced game*, which differs from the original game only in the set of allowed strategy profiles. It turns out that in the game induced by the Ultimatum Game, the proposer can not make proposals below some $x\% > 0$ of the initial budget and the recipient can no longer reject. That is, the induced game is exactly the variant of the Dictator Game where the recipient is initially endowed with $x\%$ of the budget, and the proposer has to decide whether to make an additional donation to the recipient from the remaining part of the budget. The parameter x represents the level of players’ cooperation and their perception of money. Following similar techniques to those used in [7] to compute this parameter, we obtain $x = 25\%$. In other words, cooperative equilibrium predicts that, on average, the proposer’s behaviour in the Ultimatum Game is equivalent to the dictator’s behaviour in the variant of the Dictator Game where the dictator is given 75% of the original budget and the recipient has 25% of the budget. That is, the offer in the Ultimatum Game should be equal to the donation in this variant of the Dictator Game, augmented by 25% of the budget initially given to the recipient.

To test this prediction, we conducted a behavioural experiment via Amazon Mechanical Turk [35, 26, 38]. We recruited US subjects, who were randomly assigned to two different treatments. In the first treatment, each participant was given a role of either the proposer or the recipient in the Ultimatum Game with a stake of \$0.40. The proposers were then asked to play the following variant of the Dictator Game: they were given additional \$0.30 (i.e., 75% of the budget in the Ultimatum Game) and paired with another anonymous MTurk worker who was given only \$0.10 (the remaining 25%). In the second treatment, the games were played in the reversed order.

As our results demonstrate, cooperative equilibrium correctly predicts the majority of human behaviours in such settings. Specifically, we observed that about 80% of our experimental subjects followed the cooperative equilibrium prediction. Interestingly, the remaining 20% of the subjects who deviated from the prediction were those who behaved completely inconsistently in the two games: that is, they chose to split the budget equally in the Ultimatum Game, but kept everything to themselves in the Dictator Game. We note that this type of subject appears problematic for the inequity aversion models as well, and we believe, deserves a separate consideration (see Section 6 for a thorough discussion).

Finally, this paper adds to the experimental evidence for a positive correlation between proposals in the Ultimatum Game and donations in the Dictator Game. To date, we are only aware of two published experimental studies where participants played both of these games, and one of them [36] reported a significant positive correlation between the strategies in the Ultimatum Game and the Dictator Game.² In this context, our work contributes to formally explain such a correlation between the two games through our newly defined concept of cooperative equilibrium for Pareto solvable games.

The paper unfolds as follows. We start with describing the games in consideration in Section 2. Then, in Section 3, we define our model. In detail, we first formalise our notion of fairness and then determine which strategies the players would consider feasible in this respect. We then proceed and introduce the class of Pareto solvable games and formally define a cooperative equilibrium for these games. Our theoretical results are presented in Section 4, where we compute cooperative equilibria for the Ultimatum Game and the Nash Bargaining Problem. Specifically, we show that these games are indeed Pareto solvable and have a unique cooperative equilibrium, in which both players play pure strategies. Further, Section 5 presents the experiment we conducted to test our predictions about the behaviour in the Ultimatum Game and its correlation with the behaviour in the Dictator Game. We conclude with a discussion of our results and future directions in Section 6. Additionally, the Appendix provides full instructions for our experimental subjects.

2 Preliminaries

We start with the definitions of games considered in this paper. Our interest in these game classes is motivated by the two following reasons. First, they model situations of primary importance, such as negotiation and coordination problems. Second, they provide empirical evidence of behaviours, which are inconsistent with traditional game-theoretic solutions. In particular, the Ultimatum Game that we use as our main example, has long been in the centre of attention in experimental economics.

Ultimatum Game (UG). In this game, the first player (the *proposer*) is endowed with a sum of money (or, *budget*), b , and needs to propose how to divide this money between himself and the other player (the *recipient*). That is, he chooses the amount $p(b) \in [0, b]$ to propose to the recipient. The recipient gets to decide whether to accept or reject the proposal. If he accepts, the money is split according to the proposal: that is, the recipient gets $p(b)$, and the proposer keeps the remaining $b - p(b)$. If the recipient rejects, both players end up with no money at all. We denote this game by $UG(b)$.

Dictator Game (DG). In this game, similarly to the Ultimatum Game, the first player (the *dictator*) determines how to split a budget b between himself and the second player. The second player (the *recipient*) simply accepts the donation made by the dictator. Thus, the recipient's role is entirely passive and has no effect on the outcome of the game. We denote this game by $DG(b)$.

For our study, it is also convenient to define a modified version of the Dictator Game, where the recipient may have some initial budget, b_R , and the dictator has to decide, how much (if at all) to donate to the recipient from his own budget, b_D , in addition to the recipient's initial budget. Thus, the dictator chooses a strategy $d(b_D, b_R) \in [0, b_D]$, and the recipient has no say. We denote this game by $DG(b_D, b_R)$.

Nash Bargaining Problem (NBP). This is a two-player game used to model bargaining interactions. In this game, two players demand a portion of a given budget, b . If the total amount requested by the players is less than b , they both get their request. If the total request exceeds b , both get nothing.

Coordination Game (CG). In this game, each player has to select one of the two strategies, A or B . Both players prefer any coordinated profile (A, A) or (B, B) to any uncoordinated profile, (A, B) or (B, A) . However, the players' individual utilities from playing (A, A) or (B, B) may be different.

Note also, that (A, A) and (B, B) are pure strategy Nash equilibria. Now, any probability distribution over these two equilibria is a correlated equilibrium, and the players have to decide which of the infinitely many possible such equilibria should be played. If they disagree and choose different distributions, they are likely to receive zero payoffs, while all bargaining solutions define correlated equilibria.

²In particular, [17] did not directly report any correlation between the strategies in the two games. However, upon our request, one of the authors (Antonio Espín) have confirmed that they also observe such a correlation

3 Cooperative equilibrium for Pareto solvable games

In this section, we extend the definition of cooperative equilibrium to a larger class of single-shot n -player games, which we term Pareto solvable. This class, in particular, contains the Ultimatum Game and the Nash Bargaining Problem (and hence, the Coordination Game) as above.

Our key idea is that players would divide into coalitions and act cooperatively as long as they believe that the resulting outcome of the game is “fair” to the members of each coalition. In Section 3.1, we formalise our notion of fairness for a given coalition of players, and then determine which strategies the players would consider feasible in this respect, given a particular coalition structure, that is, a partition of players into coalitions. We then proceed and introduce the class of Pareto solvable games, and formally define a cooperative equilibrium for these games in Section 3.2.

3.1 Undominated fair strategies and games generated by coalition structures

Let $\mathcal{G} = (N, (S_i, \pi_i)_{i \in N})$ be a normal-form game with a set N of n players, and, for all $i \in N$, a finite set of strategies S_i and a monetary payoff function $\pi_i : S \rightarrow \mathbb{R}$, where $S = \times_{j \in N} S_j$. For any given coalition (i.e., a subset) of players $C \subseteq N$, we write $S_C = \times_{j \in C} S_j$ for the set of strategy profiles restricted to C , and we use $-C$ to denote its complement, $N \setminus C$. Similarly, for a single player i , we use $-i$ for the set $N \setminus \{i\}$ of all players but i .

We denote by $\Delta(X)$ the set of probability distributions on a finite set X . Thus, $\Delta(S_i)$ defines the set of mixed strategies for player $i \in N$, and his expected payoff from a mixed strategy profile σ is given by $\pi_i(\sigma) = \sum_{s \in S} \pi_i(s) \sigma_1(s_1) \cdot \dots \cdot \sigma_n(s_n)$, where $s = (s_1, \dots, s_n)$.

Note that while the function π_i defines a monetary payoff of player i in game \mathcal{G} , different players may have different perceptions of the same amount of money they gain from the game. Formally, we express this idea using an *aggravation* function $\text{Agg}_i(x, y)$, defined for $x \geq y$, which represents the aggravation that player i experiences when renouncing a monetary payoff of x and receiving a monetary payoff of y . We assume the following:

- $\text{Agg}_i(x, y)$ is nonnegative:
 - $\text{Agg}_i(x, y) \geq 0$, for all $x \geq y$, and
 - $\text{Agg}_i(x, y) = 0$ if and only if $x = y$;
- $\text{Agg}_i(x, y)$ is continuous;
- $\text{Agg}_i(x, y)$ is strictly increasing in x and strictly decreasing in y .

Example 3.1. An intuitive example of an aggravation function is given by $\text{Agg}_i(x, y) = u_i(x) - u_i(y)$, where $u_i(k)$ stands for the utility of player i from getting a monetary payoff of k . It is often assumed (see [30], p. 278) that u_i is concave for positive payoffs (gains) and convex for negative payoffs (losses). That is, the same increase in the gains is perceived more significant for small amounts and less significant as the gains grow, and the other way around for the losses. In low-stake experiments, however, this effect is likely to be negligible, so one may find the simple aggravation function $\text{Agg}_i(x, y) = x - y$ practically useful, since it is parameter-free and easy to compute.

Now, let $p = (p_1, \dots, p_k)$ denote a partition of the player set N , defining a possible coalition structure that the players can form in \mathcal{G} . We denote the set of all possible partitions of N by $\mathcal{P}(N)$. Each element p_α of this partition is interpreted as a coalition: that is, the players in the same p_α play together. It is natural to expect that the players in a coalition p_α would only accept those outcomes, which pay the members of the coalition in a fair way (at the very least). With the aggravation functions as above, we define a notion of fairness as follows.

Let p_α be a coalition in the coalition structure p and s be a profile of pure strategies. Let $W_{p_\alpha}(s) = \sum_{i \in p_\alpha} \pi_i(s)$ denote the social welfare of coalition p_α , and set:

$$M_{p_\alpha}(s) = \max_{i \in p_\alpha} \text{Agg}_i(W_{p_\alpha}(s), \pi_i(s)), \quad (1)$$

$$m_{p_\alpha}(s) = \min_{i \in p_\alpha} \text{Agg}_i(W_{p_\alpha}(s), \pi_i(s)). \quad (2)$$

That is, $M_{p_\alpha}(s)$ (resp., $m_{p_\alpha}(s)$) denotes the maximum (resp., the minimum) aggravation that a player in coalition p_α has from getting his share of payoff, $\pi_i(s)$, compared to the total coalitional payoff, $W_{p_\alpha}(s)$. Intuitively, the smaller the difference between $M_{p_\alpha}(s)$ and $m_{p_\alpha}(s)$, the more fair the distribution of payoffs among the coalition members. Hence, we make the following definition.

Definition 3.2. A profile of pure strategies s is said to be *fair* for a given coalition $p_\alpha \in p$, if it minimises the difference

$$M_{p_\alpha}(s) - m_{p_\alpha}(s). \quad (3)$$

Obviously, there may be multiple fair profiles of strategies, resulting in different individual payoffs to the members of a coalition (cf. Example 3.5). We order the set of fair strategies using the super-dominance relation, introduced independently in [8] and [24].

Definition 3.3. A profile of pure strategies $s'_{p_\alpha} \in S_{p_\alpha}$ for coalition p_α *super-dominates* a profile $s_{p_\alpha} \in S_{p_\alpha}$, if for all $s_{-p_\alpha}, s'_{-p_\alpha} \in S_{-p_\alpha}$ one has $\pi_i(s'_{p_\alpha}, s'_{-p_\alpha}) > \pi_i(s_{p_\alpha}, s_{-p_\alpha})$, $\forall i \in p_\alpha$.

In words, a strategy is super-dominated by another strategy if the largest payoff that can be obtained by playing the former strategy is strictly smaller than the smallest payoff that can be obtained by playing the latter.

Now, let $\text{Fair}(\mathcal{G}, p_\alpha)$ be the set of pure strategies that are fair for coalition p_α and let $\text{Max}(\text{Fair}(\mathcal{G}, p_\alpha))$ contain its maximal elements w.r.t. the partial order induced by the super-dominance relation³. These will be named *undominated fair strategies* for coalition p_α .

Given this, we next say that a coalition structure p generates a new game, \mathcal{G}_p , where players within each coalition $p_\alpha \in p$ cooperate to maximise their total payoff, playing fair, undominated strategies. Formally,

Definition 3.4. Given a game \mathcal{G} and a coalition structure p over N , let \mathcal{G}_p be the game *generated by* p as follows. The player set in \mathcal{G}_p corresponds to the set of coalitions $p_\alpha \in p$, for each of which its set of mixed strategies is given by the closure of the convex hull of $\text{Max}(\text{Fair}(\mathcal{G}, p_\alpha))$ and the payoffs are defined as the sums of payoffs of the members of p_α .

Let p_s denote what we call the (fully) *selfish* coalition structure where each player acts individually—i.e., $p_s = (\{1\}, \dots, \{n\})$. Since every strategy is fair in such a structure, the game \mathcal{G}_{p_s} is then obtained from \mathcal{G} by simply deleting all super-dominated strategies. In general, however, \mathcal{G}_p may appear very different from \mathcal{G} and, importantly, it need *not* be the game generated by the strategies that maximise the total welfare. In particular, this may happen for the (fully) *cooperative* coalition structure $p_c = (N)$ where all the players act collectively, forming the *grand* coalition, $N = \{1, \dots, n\}$. See the following Example 3.5 for illustration.

Example 3.5. Let \mathcal{G} be the following variant of the Prisoner's dilemma:

	C	D
C	10, 10	1, 20
D	20, 1	2, 2

Assume that players have the same perception of money, which we denote Agg . Since Agg is strictly decreasing in the variable y , one can easily verify that for the cooperative coalition structure $p_c = (\{1, 2\})$, the only fair strategies are (C, C) and (D, D) . Indeed,

$$M_{\{1,2\}}(C, D) - m_{\{1,2\}}(C, D) = M_{\{1,2\}}(D, C) - m_{\{1,2\}}(D, C) = \text{Agg}(21, 1) - \text{Agg}(21, 20) > 0,$$

while

$$\begin{aligned} M_{\{1,2\}}(C, C) - m_{\{1,2\}}(C, C) &= \text{Agg}(20, 10) - \text{Agg}(20, 10) \\ &= M_{\{1,2\}}(D, D) - m_{\{1,2\}}(D, D) = \text{Agg}(4, 2) - \text{Agg}(4, 2) = 0. \end{aligned}$$

Now, note that (D, D) is super-dominated by (C, C) , which is then the only maximal fair strategy, and hence, \mathcal{G}_{p_c} is a one-player, one-strategy game. Yet, the only strategy (i.e., (C, C)) admissible in this game does *not* maximise the social welfare.

³Since strategy sets are finite, $\text{Fair}(\mathcal{G}, p_\alpha)$ always has a maximal element. Moreover, for the Ultimatum Game and the Bargaining problem, we will show that $\text{Fair}(\mathcal{G}, p_\alpha)$ is always a singleton, thus implying $\text{Max}(\text{Fair}(\mathcal{G}, p_\alpha)) = \text{Fair}(\mathcal{G}, p_\alpha)$. For these games the assumption of the strategy sets' finiteness can then be removed and, in fact, for technical convenience, we will assume these sets are continuous.

3.2 Pareto solvable games and cooperative equilibrium

Having defined the games generated by coalition structures, we now use them to formalise our extended notion of cooperative equilibrium. For simplicity of exposition, we restrict the definition to the case of two-player games, which in particular contain the Ultimatum Game and the Bargaining problem, which are the focus of this paper. We later discuss how this definition generalises to games with n players.

Consider a two-player game \mathcal{G} and fix a coalition structure p . For the game \mathcal{G}_p generated by p , let $\text{Nash}(\mathcal{G}_p)$ denote the set of its mixed strategy Nash equilibria. We then say that player i *plays according* to the coalition structure p , if he plays his corresponding marginal strategy σ_i of some Nash equilibrium profile $\sigma \in \text{Nash}(\mathcal{G}_p)$, given that other players play σ_{-i} . We say that player i *deviates* from p if he applies a unilateral profitable deviation from an equilibrium profile $\sigma \in \text{Nash}(\mathcal{G}_p)$ —that is, he plays a strategy σ'_i against σ_{-i} , for some $\sigma \in \text{Nash}(\mathcal{G}_p)$, such that $\pi_i(\sigma'_i, \sigma_{-i}) > \pi_i(\sigma)$. Note that while no player can deviate from the selfish coalition structure $p_s = (\{1\}, \{2\})$ by definition, deviations may be applied from the cooperative structure $p_c = (\{1, 2\})$.

Given a coalition structure p , a player i and a subset $D \subseteq N \setminus \{i\}$ of other players, let $\tau_D^i(p) \in [0, 1]$ denote the subjective probability that player i assigns to the event that the set of players D deviates from p for the sake of their individual interests, knowing that player i plays according to the coalition structure p . In a two-player game, we only have $D = -i$ or $D = \emptyset$, and so $\tau_{-i}^i(p) = 1 - \tau_\emptyset^i(p)$. In [7], the probabilities $\tau_D^i(p)$ were computed directly using equation (7) below. However, in light of the more recent developments, where these probabilities have been used to show the consistency of the cooperative equilibrium theory with the Social Heuristics Hypothesis [37] or predicting human behaviour in iterated social dilemmas [12], here we prefer to treat them as parameters, so that the following properties hold:

- (A1) If player $-i$ has no profitable deviation from any $\sigma \in \text{Nash}(\mathcal{G}_p)$, then $\tau_{-i}^i(p) = 0$;
- (A2) If for all $\sigma \in \text{Nash}(\mathcal{G}_p)$ and for all deviations σ'_{-i} of player $-i$ from σ there is a strategy σ'_i such that $\pi_{-i}(\sigma'_i, \sigma'_{-i}) < \pi_{-i}(\sigma)$, then $\tau_D^i(p) < 1$.

That is, if there is no incentive for a player to deviate from any equilibrium profile under a given coalition structure, then the probability of deviating from the structure is zero; and, if deviating from a coalition structure may lead to an eventual loss, then the probability of deviating is less than 1. In particular, (A1) implies that $\tau_{-i}^i(p_s) = 0$, since the sets of Nash equilibria of \mathcal{G} and \mathcal{G}_{p_s} coincide, and hence no unilateral deviations are profitable.

Now, let $e_\emptyset^i(p)$ be the minimal payoff that player i can get if nobody abandons the coalition structure p (i.e., $e_\emptyset^i(p) = \min_{\sigma \in \text{Nash}(\mathcal{G}_p)} \pi_i(\sigma)$), and let $e_{-i}^i(p)$ be the minimal payoff of player i when he plays according to some Nash equilibrium of \mathcal{G}_p and $-i$ deviates from that equilibrium profile. We define the *value* of a coalition structure p for player i as an expected value of these minimal payoffs, that is,

$$v_i(p) = e_\emptyset^i(p)\tau_\emptyset^i(p) + e_{-i}^i(p)\tau_{-i}^i(p). \quad (4)$$

In what follows, we consider a game \mathcal{G} as a tuple $(N, (S_i, \pi_i)_{i \in N}, \tau_D^i(p)_{i \in N, D \subseteq N \setminus \{i\}, p \in \mathcal{P}(N)})$, where $(N, (S_i, \pi_i)_{i \in N})$ corresponds to the original strategic game form and $\tau_D^i(p) \in [0, 1]$ are given parameters satisfying the properties (A1) and (A2).

With this notation, we now define the class of *Pareto solvable* games, as follows.

Definition 3.6. A two-player game \mathcal{G} is called *Pareto solvable* if there exists a coalition structure p^* that maximises the values $v_1(p)$ and $v_2(p)$ for both players at once.

The class of Pareto solvable games is quite large. Indeed, under the assumption of the deviation probability parameters $\tau_D^i(p)$ being symmetric (that is, $\tau_2^1(p) = \tau_1^2(p)$ for $p = p_s, p_c$), which is implied, in particular, by the formulas for their computation proposed in [7] and [12], it clearly contains all major two-player social dilemmas, such as the Prisoner's dilemma, the Traveler's dilemma and the Bertrand competition. Moreover, regardless of the exact values of $\tau_D^i(p)$'s, as long as they satisfy the properties (A1) and (A2), it also includes the Ultimatum Game and the Nash Bargaining Problem (and hence, the Coordination game), as we demonstrate in Section 4 below.

For these games, we now define the notion of cooperative equilibrium. We use the values $v_i(p^*)$ to represent the players' beliefs about the payoffs they expect to obtain in the game. In other words, these values make a tacit binding among the players, such that they only play those strategies that guarantee them payoffs of at least $v_i(p^*)$.

Definition 3.7. Let \mathcal{G} be a Pareto solvable game and p^* a coalition structure that maximises the values $v_i(p)$ for both players. The *induced* game $\text{Ind}(\mathcal{G}, p^*)$ is the restriction of \mathcal{G} where the set of (mixed) strategies is reduced to include only those profiles $\sigma \in \times_{i \in N} \Delta(S_i)$, for which $\pi_i(\sigma) \geq v_i(p^*)$ for all $i \in N$.

Finally, since the set of strategy profiles in the induced game is non-empty, convex and compact, it possesses a Nash equilibrium, and we can make the following definition.

Definition 3.8. A *cooperative equilibrium* of a Pareto solvable game \mathcal{G} is defined as a Nash equilibrium of the induced game $\text{Ind}(\mathcal{G}, p^*)$.

Remark 3.9. For the sake of simplicity, we have defined the cooperative equilibrium only for two-player Pareto solvable games. The extension to n -player games is straightforward. For example, the value for player i of a coalition structure p will be defined as

$$v_i(p) = \sum_{C \subseteq N \setminus \{i\}} e_C^i(p) \tau_C^i(p),$$

where $\tau_C^i(p)$ is the subjective probability that player i assigns to the event that players in C abandon the coalition structure p and $e_C^i(p)$ is the minimum payoff that player i would get if he plays according to the coalition structure p and players in C abandon it. Given this definition, one may define n -player Pareto solvable games as those games for which there is a coalition structure p^* such that $v_i(p)$ is maximised, for all $i = 1, \dots, n$. In a similar fashion, all other definitions extend to n player games.

4 Cooperative equilibria of Ultimatum Games and Nash Bargaining Problems

In this section, we compute cooperative equilibria for some important subclasses of Pareto solvable games: namely, for the Ultimatum Game and the Nash Bargaining Problem. Specifically, we demonstrate that these games are indeed Pareto solvable and have a unique cooperative equilibrium in pure strategies.

4.1 Purifiable games

The existence of a pure strategy cooperative equilibrium is due to the fact that the induced games of UG and NBP are what we term *purifiable*, in the sense that for every mixed strategy σ_i of any player i , there exists a pure strategy s_i , which is a profitable deviation from σ_i , whichever strategy σ_{-i} is played by the other player(s). Formally,

Definition 4.1. A game \mathcal{G} is called *purifiable* if for each player $i \in N$ and any mixed strategy $\sigma_i \in \Delta(S_i)$, there is a pure strategy $s_i \in S_i$ such that $\pi_i(s_i, \sigma_{-i}) \geq \pi_i(\sigma_i, \sigma_{-i})$, for all $\sigma_{-i} \in \times_{j \neq i} \Delta(S_j)$.

For such games, it then suffices to consider only pure strategies. Now, in both the UG and the NBP models, the two players share a unit of payoff (i.e., we normalise the budget b to be equal to 1) between them – that is, the total sum of the players' payoffs is always 1. As the following Lemma 4.2 shows, there is then only one way to share the payoffs so that the players have equal aggravations.

Lemma 4.2. Let $\text{Agg}_i(x, y)$ be the aggravation function of player $i = 1, 2$ in game \mathcal{G} . Then, the equation $\text{Agg}_1(1, 1 - x) = \text{Agg}_2(1, x)$, with $x \in [0, 1]$, has a unique solution x^* . Moreover, $x^* \neq 0, 1$.

Proof. The proof is straightforward. Let $f_1(x) = \text{Agg}_1(1, 1 - x)$ and $f_2(x) = \text{Agg}_2(1, x)$. Then,

- $f_1(x)$ is a strictly increasing continuous function with $f_1(0) = 0$ and $f_1(1) > 0$; and
- $f_2(x)$ is a strictly decreasing continuous function with $f_2(0) > 0$ and $f_2(1) = 0$.

Hence, there is a unique point x^* where $f_1(x^*) = f_2(x^*)$, and $x^* \neq 0, 1$. □

We now use Lemma 4.2 to describe cooperative equilibria of the Ultimatum Game and the Nash Bargaining Problem, as follows.

4.2 Ultimatum Games

Theorem 4.3. *Let \mathcal{G} be the Ultimatum Game with players P (the proposer) and R (the recipient), and budget b . Let x^* be the solution to $\text{Agg}_P(1, 1 - x) = \text{Agg}_R(1, x)$. Then,*

- \mathcal{G} is Pareto solvable, where the values for both players are maximised by the cooperative coalition structure p_c and equal:

$$v_P(p_c) = b(1 - x^*) \quad \text{and} \quad v_R(p_c) = bx^* \tau_\emptyset^R(p_c).$$

- The induced game $\text{Ind}(\mathcal{G}, p_c)$ is purifiable, and its purification coincides with the variant of the Dictator Game with parameters $b_D = b(1 - x^* \tau_\emptyset^R(p_c))$ and $b_R = bx^* \tau_\emptyset^R(p_c)$, where the proposer plays the role of a dictator and donates to the recipient the amount of at most $bx^* \tau_P^R(p_c)$.

Proof. First, observe that the game generated by the selfish coalition structure $p_s = (\{P\}, \{R\})$ coincides with the original game (i.e., $\mathcal{G}_{p_s} = \mathcal{G}$), as there are no super-dominated strategies. Now, while this game may have different Nash equilibria, the lowest payoff for both players is obtained at the equilibrium profile $(0, R)$ where the proposer makes an offer of 0 to the recipient, and the recipient rejects. In this case, both players get nothing, and since by (A1) the probability of deviation from p_s is 0 for both players, they both have a zero value for the selfish coalition structure p_s . Thus, $v_P(p_s) = v_R(p_s) = 0$.

For the cooperative coalition structure, $p_c = (\{P, R\})$, we now demonstrate that its generated game \mathcal{G}_{p_c} only allows a single strategy profile (bx^*, A) , where the proposer makes an offer of bx^* and the recipient accepts. To this end, we only need to show that, for all $x \in [0, 1]$ such that $x \neq x^*$, one has

$$M_{\{P, R\}}(bx^*, A) - m_{\{P, R\}}(bx^*, A) < M_{\{P, R\}}(bx, A) - m_{\{P, R\}}(bx, A). \quad (5)$$

But this is straightforward from the definition of x^* as the unique solution for

$$M_{\{P, R\}}(bx, A) - m_{\{P, R\}}(bx, A) = b |\text{Agg}_P(1, 1 - x) - \text{Agg}_R(1, x)| = 0.$$

implying that the left-hand side of inequality (5) equals zero, and the right-hand side is strictly positive.

Now, if the proposer is assumed to make an offer of $bx^* > 0$, then it is not beneficial for the recipient to deviate from accepting the offer, and hence, by (A1), $\tau_\emptyset^P(p_c) = 1$. Thereby, since $e_\emptyset^P(p_c) = b(1 - x^*)$, we have $v_P(p_c) = b(1 - x^*) > 0$.

On the other hand, if it is believed that the recipient will accept, then the proposer has an incentive to deviate and make a smaller offer. In the worst case for the recipient, the proposer offers him zero, so $e_\emptyset^R(p_c) = 0$. We thus get $v_R(p_c) = bx^* \tau_\emptyset^R(p_c)$. Finally, by (A2), we have that $\tau_\emptyset^R(p_c) > 0$, which, coupled with $bx^* > 0$, yields $v_R(p_c) > 0$.

Hence, the cooperative coalition structure p_c maximises the values for both players, so the Ultimatum Game is Pareto solvable.

We now turn to prove the second part of the theorem, stating that the induced game $\text{Ind}(\mathcal{G}, p_c)$ is purifiable. To avoid unnecessary technical complications, we only consider finitely supported strategies for the proposer, even if his strategy set is given by the continuous interval $[0, b]$. That is, we assume that any proposer's mixed strategy σ_P is given by a finite sequence of coefficients $(\lambda_{(1)}, \dots, \lambda_{(K)})$, for some $K \in \mathbb{N}$, such that $\lambda_{(k)} > 0$ represents the probability of playing a pure strategy $s_{(k)} \in [0, b]$. Hence, $\sum_{k=1}^K \lambda_{(k)} = 1$, and w.l.o.g. we assume that the strategies $s_{(k)}$ are ordered in ascending order.

We next demonstrate that any such mixed strategy σ_P can be replaced by the pure strategy $s_{(1)}$. Indeed, for any mixed strategy $\sigma_R = (\lambda, 1 - \lambda)$ of the recipient, with λ being the probability of $s_R = A$ (accept), we get

$$\begin{aligned} \pi_P(\sigma_P, \sigma_R) &= \lambda (\lambda_{(1)}(b - s_{(1)}) + \dots + \lambda_{(K)}(b - s_{(K)})) \\ &< \lambda(b - s_{(1)})(\lambda_{(1)} + \dots + \lambda_{(K)}) \\ &= \lambda(b - s_{(1)}) = \pi_P(s_{(1)}, \sigma_R). \end{aligned}$$

Similarly, any mixed strategy $\sigma_R = (\lambda, 1 - \lambda)$ of the recipient, where $\lambda < 1$, can be replaced by the pure strategy $s_R = A$, as for any strategy σ_P of the proposer we have

$$\begin{aligned} \pi_R(\sigma_P, \sigma_R) &= \lambda (\lambda_{(1)}s_{(1)} + \dots + \lambda_{(K)}s_{(K)}) \\ &< \lambda_{(1)}s_{(1)} + \dots + \lambda_{(K)}s_{(K)} \\ &= \pi_R(\sigma_P, A). \end{aligned}$$

Finally, in the purification of the induced game $\text{Ind}(\mathcal{G}, p_c)$, the proposer makes an offer of at least $v_R(p_c) = bx^*\tau_\emptyset^R(p_c)$, but of at most $b - v_P(p_c) = bx^*$, and the recipient must accept. The game then coincides with the variant of the Dictator Game with parameters $b_P = b(1 - x^*\tau_\emptyset^R(p_c)) \geq b(1 - x^*)$ and $b_R = bx^*\tau_\emptyset^R(p_c)$, where the donation that the dictator can make to the recipient does not exceed

$$\begin{aligned} b_P - v_P(p_c) &= b(1 - x^*\tau_\emptyset^R(p_c) - (1 - x^*)) \\ &= bx^*(1 - \tau_\emptyset^R(p_c)) = bx^*\tau_P^R(p_c). \end{aligned}$$

This completes the proof. \square

Theorem 4.3 implies that the Ultimatum Game has a unique pure strategy cooperative equilibrium, in which the proposer makes an offer of $x^*\tau_\emptyset^R(p_c)$, and the recipient accepts. Moreover, since the induced game of UG essentially coincides with a corresponding variant of the Dictator Game, the cooperative equilibrium predicts a correlation between these two games. Specifically, the theorem suggests that

Corollary 4.4. *In a Ultimatum Game \mathcal{G} , the proposer makes an offer of*

$$bx^*\tau_\emptyset^R(p_c) + d(b(1 - x^*\tau_\emptyset^R(p_c)), bx^*\tau_\emptyset^R(p_c)), \quad (6)$$

where $d(b(1 - x^*\tau_\emptyset^R(p_c)), bx^*\tau_\emptyset^R(p_c))$ is the donation that the proposer would make to the recipient if he were to act as the dictator in $DG(b_P, b_R)$ with parameters $b_P = b(1 - x^*\tau_\emptyset^R(p_c))$ and $b_R = bx^*\tau_\emptyset^R(p_c)$, and we expect that $d(b(1 - x^*\tau_\emptyset^R(p_c)), bx^*\tau_\emptyset^R(p_c)) \leq bx^*\tau_P^R(p_c)$.

Now, in order to obtain testable predictions of players' behaviour, we need to evaluate the parameters x^* and $\tau_D^i(p)$. As it is often in experiments with anonymous participants, the average experimental subjects may overgeneralise their experience and expect their opponents to have the same perception of money as they do, so, in order to make predictions of average behaviour in anonymous games, we assume that all players have the same aggravation function. That is, for any two subjects i and j , we have $\text{Agg}_i(x, y) = \text{Agg}_j(x, y)$. Applying this assumption in Lemma 4.2, we get $x^* = 0.5$, which allows us to make the following individual level predictions:

- *Local prediction.* In a UG experiment with anonymous players, *most* proposers will offer to their opponents a payment of $b \cdot 0.5\tau_\emptyset^R(p_c) + d(b(1 - 0.5\tau_\emptyset^R(p_c)), b \cdot 0.5\tau_\emptyset^R(p_c))$.

To estimate the probabilities of deviation from a given coalition structure, we adopt the idea from [7]. Specifically, these probabilities depend on the *incentive* for each player to abandon a coalition structure, and his possible *loss* from doing so. Here, we define the incentive, $I_{-i}(p)$, for a player $-i \in N$ to abandon the coalition structure p , as the aggravation in renouncing to the maximal possible gain that $-i$ can obtain by leaving the coalition structure p , while his opponent (player i) plays according to p :

$$I_{-i}(p) = \sup \{ \text{Agg}_{-i}(\pi_{-i}(\sigma_i^p, \sigma_{-i}), \pi_{-i}(\sigma_i^p, \sigma_{-i}^p)) : \sigma^p \in \text{Nash}(\mathcal{G}_p), \sigma_{-i} \in \Delta(S_{-i}) \}$$

However, player i may also deviate from p (whether he anticipates player $-i$'s deviation or not), thus yielding a loss for player $-i$. We define such a loss, $L_{-i}(p)$, of a player $-i$ when he leaves a coalition structure p in order to maximise his payoff as the maximal aggravation that he may experience, while player i also may deviate from the coalition structure p (either following his selfish interests or anticipating player $-i$'s deviation). Formally,

$$L_{-i}(p) = \sup \left\{ \text{Agg}_{-i}(\pi_{-i}(\sigma_i, \sigma_{-i}), \pi_{-i}(\sigma_i^p, \sigma_{-i}^p)) : \begin{array}{l} \sigma^p \in \text{Nash}(\mathcal{G}_p), \sigma_{-i} \in \arg\max_{\Delta(S_{-i})} \pi_{-i}(\sigma_i^p, \sigma_{-i}), \sigma_i \in \Delta(S_i) \\ \text{s.t. } \pi_i(\sigma_i, \sigma_{-i}^p) > \pi_i(\sigma_i^p, \sigma_{-i}^p) \quad \text{or} \quad \pi_i(\sigma_i, \sigma_{-i}) > \pi_i(\sigma_i^p, \sigma_{-i}) \end{array} \right\}$$

That is, for each $\sigma^p \in \text{Nash}(\mathcal{G}_p)$, mixed strategy σ_{-i} maximises the payoff of player $-i$ from $(\sigma_i^p, \sigma_{-i})$, and σ_i runs over the set of i -deviations from either $(\sigma_i^p, \sigma_{-i}^p)$ or $(\sigma_i^p, \sigma_{-i})$.

Given this, we finally set:

$$\tau_{-i}^i(p) = \frac{I_{-i}(p)}{I_{-i}(p) + L_{-i}(p)} \quad (7)$$

As we previously mentioned in Example 3.1, in a low-stake experiment, function $\text{Agg}_i(x, y) = x - y$ appears a good approximation of the real players' aggravation. For such a function, it is immediate that $I_{-i}(p_c) = x^* = 0.5$ and $L_{-i}(p_c) = 1 - x^* = 0.5$. We therefore can make the following prediction:

- *Global prediction.* In a UG experiment with anonymous players and low stakes, it is expected that the average offer is composed of 25% of the total endowment, b , and the average donation in DG(0.75 b , 0.25 b).

We test these predictions experimentally in Section 5. We conclude the section with presenting analogous results for the Nash Bargaining Problem.

4.3 Nash Bargaining Problem

Consider the classical Nash Bargaining Problem, where two players bargain on how to share b units of surplus, where they get their claims satisfied if and only if these claims are feasible, that is, they sum up to less than or equal to b . Then,

Theorem 4.5. *In the only pure cooperative equilibrium of the Nash Bargaining Problem, one player asks for bx^* and the other player claims $b(1 - x^*)$.*

The proof of Theorem 4.5 follows similar lines to the proof of Theorem 4.3, and hence is omitted. Finally, since for players with the same perception of money we have $x^* = 0.5$, this yields the following corollary.

Corollary 4.6. *If the players have the same perception of money, then in the unique cooperative equilibrium of the Nash Bargaining Problem, each player asks for exactly half of the endowment.*

5 Behavioural experiment investigating the correlation between Ultimatum Game and Dictator Game

In this section, we report on the experiment we conducted to test the correlation between the UG offers and the DG donations as our model predicts. We start with the outline of the experiment, and then discuss the results we obtained.

5.1 Design

We recruited US subjects (a.k.a. *workers*) via Amazon Mechanical Turk (AMT) [35, 26, 38]. As is common in experiments with human participants (either online or in physical laboratories), AMT workers receive a baseline payment for participation and can also earn an additional bonus that depends on how they perform in the game. The main advantage of the AMT experiments is in that they are easily implementable and relatively inexpensive, compared to laboratory experiments. Nevertheless, it has been shown that data gathered using AMT agree both qualitatively and quantitatively with data collected in physical labs [26, 38, 42].

However, when planning such an experiment, one should be aware of the fact that AMT workers may try to play the same game multiple times in order to get a larger bonus and/or select random strategies in order to minimise the time spent on completing the task. To control for these issues, we used Qualtrics—a survey builder that allowed us to exclude those workers who have already participated in the experiment (by checking their AMT ID and their IP address) or those who failed to answer the comprehension questions that we designed for each part of the task to make sure that the workers understand the task and would make sensible decisions.

The task that the workers had to perform in the experiment consisted of two parts: play a given role in the Ultimatum Game and in the Dictator Game. The experiment contained two treatments, differing in the order in which the UG and the DG were played. In both treatments, subjects earned \$0.20 for participating and received a bonus that was determined from their performance in the two games. In more detail:

- *Treatment 1.* Subjects were randomly assigned a role of either the proposer/dictator or the recipient. First, they played the UG as follows. The proposers were given a budget of \$0.40 and asked how much, if at all, they would offer to an anonymous recipient they have been randomly paired with, while the recipient may either accept or reject the offer. Responders, at the same time, were asked

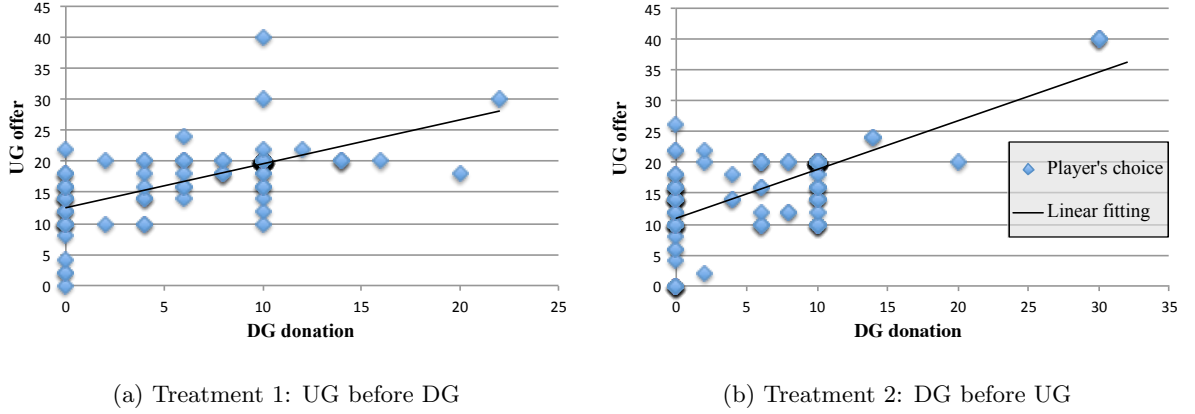


Figure 1: Scatter plot of the players' choices in the two treatments: UG played after DG (a) and DG played after UG (b), after filtering the players with 0c donation in DG and an equal split offer in UG.

to declare their minimal acceptable offer.

Second, the DG was played. The proposers (playing now the role of a dictator) were notified that they have been assigned another anonymous worker to interact with. They were given another \$0.30 and informed that their current partner is endowed with only \$0.10. Then they were asked how much, if at all, they would donate to the other participant, who has no say and can only accept their offer.

- *Treatment 2.* As above, but playing the DG first and the UG second.

In both games, we restricted the set of possible strategies to multiples of 2 cents, within the given budget. Thus, a proposer in the Ultimatum Game could only make an offer of $\{0, 2, 4, \dots, 40\}$ cents. Full instructions for participants are presented in the Appendix.

5.2 Results

The results for Treatment 1 are summarised in Table 1 and Table 2. We had 172 subjects that played as proposers in the Ultimatum Game and as dictators in the Dictator Game.

In detail, the mean offer in the UG was 18.18c, which constitutes 45% of the budget. More than half of the subjects (about 65%) offered the equal split, which was also the modal and the median offer. Only one subject acted consistently with the subgame perfect equilibrium for payoff-maximising preferences, which is to offer the recipient nothing at all. We note that these results appear consistent with those reported in other studies [28, 6, 34].

In the DG, the mean offer was 5.40c. The distribution of donations was essentially bimodal: 42% of the subjects decided to act selfishly and make no donation at all, while 40% of the participants acted so as to minimise inequity and donate 10c. These results are qualitatively similar to those presented in previous studies, although in previous reports the selfish behaviour was observed significantly more often than the inequity averse behaviour [15]. This difference may appear related to the framing effect caused by the fact that the DG was played after the UG in this treatment. However, we cannot reject the null hypothesis that donations in Treatment 1 and Treatment 2 (differing by the order of the UG and DG) are actually the same. In fact, we ran the non-parametric Wilcoxon Rank-sum test, i.e., the standard hypothesis test for data samples that cannot assumed to be normally distributed like our case [45], and the returned significance value is $P = 0.1969$, which does not reject the null hypothesis.

In contrast, we find a significant positive correlation between the UG offers and the DG donations (Spearman's ρ rank correlation test: $\rho = 0.3412, P < .0001$). In particular, we observe that 43% of the subjects played exactly according to our global prediction based on the cooperative equilibrium:

$$p(40) = d(30, 10) + 10 \quad (8)$$

That is, their offer in the Ultimatum Game with $b = 40c$ was composed of their donation in the Dictator Game with parameters $b_D = 0.75b = 30$ and $b_R = 0.25b = 10$, and additional $10c$ (i.e., 25% of the total endowment in the UG).

We test the quality of this prediction by fitting the following linear equation:

$$p(40) = c_{\text{fitting}}(1)d(30, 10) + c_{\text{fitting}}(0) \quad (9)$$

where the coefficients $c_{\text{fitting}}(0)$, $c_{\text{fitting}}(1)$ are estimated through least square error. This linear fit is also showed in Figure 1a. We compare the coefficients with the coefficients $c_{\text{model}}(0)$ and $c_{\text{model}}(1)$ fitted by our prediction model by computing their corresponding mean squared errors, $\text{MSE}_{\text{fitting}}$ and $\text{MSE}_{\text{model}}$. We find that $c_{\text{fitting}}(0) = 16.3912$ and $c_{\text{fitting}}(1) = 0.3320$, with the mean squared error $\text{MSE}_{\text{fitting}} = 17.6383$, while the cooperative equilibrium model prescribes $c_{\text{model}}(0) = 10$ and $c_{\text{model}}(1) = 1$, resulting in the mean squared error $\text{MSE}_{\text{model}} = 37.0467$.

At first glance, our prediction may seem quite inaccurate, compared to the linear fitting. However, it turns out that the difference in equations (8) and (9) is mainly due to a subset of 42 subjects (about 24%) who demonstrated completely inconsistent behaviour in playing the two games—that is, they decided for the equal split in the UG (offering the recipient $20c$), but acted selfishly in the DG (donating $0c$). Note that this type of subjects is also very problematic for the inequity aversion models, as they act in an inequity averse way in the UG and the opposite in the DG. We refer to Section 6 for a more detailed discussion about this issue.

Taking out these 42 inconsistent subjects allows us to observe significantly stronger correlations. When only consistent participants were taken into account, the mean UG offer was equal $17.60c$ and the mean DG donation was $7.15c$. The correlation between the UG offers and the DG donation also gets much stronger (Spearman's ρ rank correlation: $\rho = 0.7136$, $P < .0001$). Moreover, 57% of these subjects played exactly according to equation (8)). Finally, the linear fitting now gives $p_{\text{fitting}}(0) = 12.5308$, $p_{\text{fitting}}(1) = 0.7086$ and $\text{MSE}_{\text{fitting}} = 14.6281$, which is much closer to the cooperative equilibrium model with $p_{\text{model}}(0) = 10$, $p_{\text{model}}(1) = 1$ and $\text{MSE}_{\text{model}} = 16.7077$. This trend is also shown by the scatter plot of the players' choices together with their linear fitting in Figure 1a.

	Mean Offer	Mean donation	Correlation
All	18.18	5.40	0.3412***
No (20,0)	17.60	7.15	0.7136***

Table 1: Statistical analysis of Treatment 1. Spearman's ρ rank correlation test. *** stands for $P < .0001$.

	Linear fitting	Model	MSE fitting	MSE model
All	$p(0) = 16.40, p(1) = 0.33$	$p(0) = 10, p(1) = 1$	17.64	37.05
No (20,0)	$p(0) = 12.53, p(1) = 0.71$	$p(0) = 10, p(1) = 1$	14.63	16.71

Table 2: Comparison between polynomial fitting model at degree $n = 1$ and predictions of the cooperative equilibrium in Treatment 1.

The results for Treatment 2 are given in Tables 3 and 4. We had 198 subjects that played as dictators in the Dictator Game and as proposers in the Ultimatum Game. The mean donation was $5.39c$, while the distribution of donations was essentially bimodal with half of the participants donating $0c$ and 37% making the inequity minimising donation of $10c$. While these distributions of donations seems to be more left-weighted than the one we observed in Treatment 1, as we mentioned before, we cannot exclude the null hypothesis that they are actually the same (Wilcoxon rank-sum: $P = 0.1969$). The mean UG offer was $16.77c$, that is 42% of the budget. Slightly more than half of the subjects (53%) chose the equal split and ten subjects acted consistently with the subgame perfect equilibrium for payoff-maximising preferences, which is to offer $0c$ or $2c$.

Similarly to Treatment 1, we find a significant positive correlation between the UG offers and the DG donations (Spearman's ρ rank correlation: test: $\rho = 0.3322$, $P < .0001$). In this case, 38% of the

subjects played exactly according to our global prediction stated in Equation (8). Again, we tested the quality of a linear fitting (see Equation (9)). We obtained $c_{\text{fitting}}(0) = 14.2689$ and $c_{\text{fitting}}(1) = 0.5079$ and the mean squared error $\text{MSE}_{\text{fitting}} = 35.5953$, while the cooperative equilibrium prediction gives $c_{\text{model}}(0) = 0.10$, $c_{\text{model}}(1) = 1$ and the mean squared error $\text{MSE}_{\text{model}} = 40.7273$.

Also in this case, the difference is mainly caused by the inconsistent behaviour of 42 subjects (21%) who donated 0c in the DG and proposed the equal split in the UG. When only consistent participants were considered, the mean UG offer was equal to 15.91c and the mean DG donation was 6.27c. The correlation between the UG offers and the DG donation also gets much stronger (Spearman's ρ rank correlation: $\rho = 0.6377$, $P < .0001$). This correlation is also apparent in the scatter plot and linear fitting in Figure 1b. Moreover, 49% of these subjects played exactly according to Equation (8), and the polynomial fitting gave $c_{\text{fitting}}(0) = 10.9531$, $c_{\text{fitting}}(1) = 0.7907$ and $\text{MSE}_{\text{fitting}} = 23.1418$, compared with the cooperative equilibrium prediction with $c_{\text{model}}(0) = 10$, $c_{\text{model}}(1) = 1$ and $\text{MSE}_{\text{model}} = 24.7695$.

	Mean donation	Mean offer	Correlation
All	5.39	16.77	0.3322***
No (0,20)	6.27	15.91	0.6377***

Table 3: Statistical analysis of Treatment 2. Spearman's ρ rank correlation test. *** stands for $P < .0001$.

	Linear fitting	Model	MSE fitting	MSE model
All	$p(0) = 14.27, p(1) = 0.51$	$p(0) = 10, p(1) = 1$	35.60	40.73
No (0,20)	$p(0) = 10.95, p(1) = 0.79$	$p(0) = 10, p(1) = 1$	23.14	24.77

Table 4: Comparison between polynomial fitting model at degree $n = 1$ and predictions of the cooperative equilibrium in Treatment 2.

6 Discussion

In this paper, we have extended the concept of cooperative equilibrium from two-player social dilemmas to a more general class of games containing, as important instances, the Ultimatum Game and the Nash Bargaining Problem. We used this concept to obtain theoretical predictions of human players' behaviour in these games. In particular, we derived a correlation between the offers that proposers make to recipients in the UG and the donations they would make in the corresponding variant of the Dictator Game where the proposer is endowed with 75% of the total budget, while the recipient has the remaining 25% upfront. This correlation is particularly interesting because it can be tested in the laboratory.

Our model explicitly predicts the regularities III and IV outlined in Section 1. Specifically, it suggests that on average, UG proposers would offer at least 25% of their budget to recipients, which is consistent with the experimental findings showing that there are almost no offers below $0.2b$ (regularity III). Also notice, that accepting less than 25% of the proposer's budget is excluded from the set of feasible strategy profiles in the game induced by the cooperative coalition structure, which is consistent with regularity IV, stating that offers below $0.25b$ get rejected with high probability. Finally, the correlation we established between the UG offers and the DG donations gives an *a posteriori* explanation of regularities I and II: indeed, virtually every donation in the DG is bounded above by the equal split [15].

In order to test the predicted correlation between the offers in the UG and the donations in the DG, we conducted an experiment via Amazon Mechanical Turk and found that about 80% of experimental subjects played according to our prediction. As for the remaining 20%, we must point out that these subjects seem to form a special class showing inconsistent behaviour in the two games: these subjects made a zero donation in the Dictator Game, but offered half of the budget to the recipient in the Ultimatum Game. Such behaviour appears problematic not only for our approach, but also for the standard inequity aversion model introduced by Fehr and Schmidt [19]. This model assumes that player i 's utility from the

allocation (x_i, x_{-i}) of monetary payoffs between himself and player $-i$, is given by:

$$u_i(x_i, x_{-i}) = x_i - \alpha_i \max\{x_{-i} - x_i, 0\} - \beta_i \max\{x_i - x_{-i}, 0\} \quad (10)$$

where $0 \leq \beta_i \leq \alpha_i$ are individual parameters, representing the extent to which player i regrets being worse off (parameter α_i) or enjoys being better off (parameter β_i) than the other player. By estimating the joint distribution (α, β) of the parameters, the model gives 30% of players with $(\alpha, \beta) = (0, 0)$, 30% of players with $(\alpha, \beta) = (0.5, 0.25)$, 30% with $(\alpha, \beta) = (1, 0.6)$, and 10% with $(\alpha, \beta) = (4, 0.6)$. However, this distribution is not consistent with the aforementioned subjects choosing equal split of the budget in the UG, but a zero donation in the DG. To see this, first observe that if a subject makes a zero donation in the Dictator Game, then his parameter β must be smaller than 0.5. Indeed, a subject with $\beta = 0.5$ is indifferent to any donation and a subject with $\beta > 0.5$ would donate half of the endowment. Also note that a person donating zero in the DG but splitting the budget equally in the UG cannot be of the type with $(\alpha, \beta) = (0.5, 0.25)$, since then he would donate a half of his endowment also in the DG. It follows that, according to the Fehr and Schmidt's estimation of the parameters, the only possibility is that such a person has the type with $(\alpha, \beta) = (0, 0)$. We further make the following few observations (see the Appendix for the full proof):

1. Subjects of type $(\alpha, \beta) = (0, 0)$ who believe they are playing with subjects of type $(\alpha, \beta) = (0, 0)$, would donate nothing in the DG and offer nothing in the UG.
2. Subjects of type $(\alpha, \beta) = (0, 0)$ who believe they are playing with subjects of type $(\alpha, \beta) = (0.5, 0.25)$, would donate nothing in the DG and offer 10c in the UG.
3. Subjects of type $(\alpha, \beta) = (0, 0)$ who believe they are playing with subjects of type $(\alpha, \beta) = (2, 0.6)$, would donate nothing in the DG and offer 16c in the UG.
4. Subjects of type $(\alpha, \beta) = (0, 0)$ who believe they are playing with subjects of type $(\alpha, \beta) = (4, 0.6)$, would donate nothing in the DG and 17.78c in the UG.

Now, it is clear from the above that Fehr and Schmidt's model does not account for the players offering equal split in the UG and making a zero donation in the DG. The question is though who are these subjects and why do they act in such an apparently inconsistent way? We note that, although their behaviour deviates from our global prediction that is based on an explicit computation of the model parameters, yet it is consistent with Theorem 4.3, corresponding to the case where $\tau_\emptyset^R(p_c) = 1$ and both the proposer and the recipient have the same perception of money. Recalling that $\tau_\emptyset^R(p_c)$ represents the probability that the recipient assigns to the event that the proposer does not deviate from the cooperative coalition structure, it follows that the cooperative equilibrium concept explains such subjects as more cooperative than the average. Testing this (as we see it, non-trivial) hypothesis is an interesting direction for future research.

There is also another interesting novel prediction provided by our cooperative equilibrium model. Consider the situation where it is common knowledge that the proposer is much richer than the recipient. This could be formalised by saying that the aggravation function of the proposer, for fixed y , increases much slower than that of the recipient. Consequently, the intersection point x^* found in Lemma 4.2 lies much closer to 1. Keeping $\tau_{R,P}(p_c)$ constant, this implies that the prediction of the cooperative equilibrium is that, in this situation, the proposer's offer would be higher than the one he would make if they were equally rich. Consequently, the cooperative equilibrium predicts that proposers offer more to poor players than to equally rich players. Such a behaviour has been qualitatively observed in the Dictator Game [4], but we are not aware of any study investigating it in the Ultimatum Game. In future research, it would be interesting to test this prediction experimentally.

References

- [1] Hélène Barcelo and Valerio Capraro. Group size effect on cooperation in one-shot social dilemmas. *Scientific Reports*, 5:7937, 2015.
- [2] Ken Binmore and Avner Shaked. Experimental economics: Where next? *Journal of Economic Behavior & Organization*, 73(1):87–100, 2010.

- [3] Pablo Brañas-Garza. Poverty in dictator games: Awakening solidarity. *Journal of Economic Behavior & Organization*, 60(3):306–320, 2006.
- [4] Pablo Brañas-Garza. Promoting helping behavior with framing in dictator games. *Journal of Economic Psychology*, 28(4):477–486, 2007.
- [5] Terence C Burnham. Engineering altruism: a theoretical and experimental investigation of anonymity and gift giving. *Journal of Economic Behavior & Organization*, 50(1):133–144, 2003.
- [6] Colin Camerer. *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press, 2003.
- [7] Valerio Capraro. A model of human cooperation in social dilemmas. *PLoS ONE*, 8(8):e72427, 2013.
- [8] Valerio Capraro. A solution concept for games with altruism and cooperation. *arXiv preprint arXiv:1302.3988*, 2013.
- [9] Valerio Capraro and Hélène Barcelo. Group size effect on cooperation in one-shot social dilemmas II. Curvilinear effect. *PLoS ONE*, 2015.
- [10] Valerio Capraro, Jillian J. Jordan, and David G. Rand. Heuristics guide the implementation of social preferences in one-shot prisoner’s dilemma experiments. *Scientific Reports*, 4:6970, 2014.
- [11] Valerio Capraro, Conor Smyth, Kalliopi Mylona, and Graham A. Niblo. Benevolent characteristics promote cooperative behaviour among humans. *PLoS ONE*, 9(8):e102881, 2014.
- [12] Valerio Capraro, Matteo Venanzi, Maria Polukarov, and Nicholas R. Jennings. Cooperative equilibria in iterated social dilemmas. In *Proc. Sixth International Symposium on Algorithmic Game Theory (SAGT ’13)*, pages 146–158, 2013.
- [13] Gary Charness and Uri Gneezy. What’s in a name? anonymity and social distance in dictator and ultimatum games. *Journal of Economic Behavior & Organization*, 68(1):29–35, 2008.
- [14] Gary Charness and Matthew Rabin. Understanding social preferences with simple tests. *Quarterly journal of Economics*, 117(3):817–869, 2002.
- [15] Christoph Engel. Dictator games: a meta study. *Experimental Economics*, 14(4):583–610, 2011.
- [16] Dirk Engelmann and Martin Strobel. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American economic review*, 94(4):857–869, 2004.
- [17] Filippos Exadaktylos, Antonio M Espín, and Pablo Brañas-Garza. Experimental subjects are not different. *Scientific reports*, 3, 2013.
- [18] Ernst Fehr and Urs Fischbacher. The nature of human altruism. *Nature*, 425(6960):785–791, 2003.
- [19] Ernst Fehr and Klaus M Schmidt. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114(3):817–868, 1999.
- [20] James H Fowler and Nicholas A Christakis. A random world is a fair world. *Proceedings of the National Academy of Sciences*, 110(7):2440–2441, 2013.
- [21] Axel Franzen and Sonja Pointner. The external validity of giving in the dictator game. *Experimental Economics*, 16(2):155–169, 2013.
- [22] Jacob K Goeree and Charles A Holt. Ten little treasures of game theory and ten intuitive contradictions. *American Economic Review*, 91:1402–1422, 2001.
- [23] W. Güth, R. Schmittberger, and B. Schwarze. An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization*, 3(4):367–388, 1982.
- [24] J. Y. Halpern and R. Pass. Game theory with translucent players. In *Theoretical Aspects of Rationality and Knowledge: Proc. Fourteenth Conference (TARK 2013)*, pages 216–221, 2013.

- [25] Joseph Henrich, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath. In search of homo economicus: Behavioral experiments in 15 small-scale societies. *American Economic Review*, 91(2):73–78, 2001.
- [26] John J Horton, David G Rand, and Richard J Zeckhauser. The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, 14(3):399–425, 2011.
- [27] R Mark Isaac and James M Walker. Group size effects in public goods provision: The voluntary contributions mechanism. *The Quarterly Journal of Economics*, 103(1):179–199, 1988.
- [28] John H Kagel and Alvin E Roth. *The handbook of experimental economics*. Princeton University Press, Princeton, NJ, 1995.
- [29] Daniel Kahneman, Jack L Knetsch, and Richard H Thaler. Fairness and the assumptions of economics. *Journal of Business*, 59(4):S285–S300, 1986.
- [30] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979.
- [31] Richard P Larrick and Sally Blount. The claiming effect: Why players are more generous in social dilemmas than in ultimatum games. *Journal of Personality and Social Psychology*, 72(4):810, 1997.
- [32] Haijin Lin and Shyam Sunder. Using experimental data to model bargaining behavior in ultimatum games. In *Experimental Business Research*, pages 373–397. Springer, 2002.
- [33] Martin A Nowak. Five rules for the evolution of cooperation. *Science*, 314(5805):1560–1563, 2006.
- [34] Hessel Oosterbeek, Randolph Sloof, and Gijs Van De Kuilen. Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental Economics*, 7(2):171–188, 2004.
- [35] Gabriele Paolacci, Jesse Chandler, and Panagiotis G Ipeirotis. Running experiments on Amazon Mechanical Turk. *Judgment and Decision mKing*, 5(5):411–419, 2010.
- [36] Alexander Peysakhovich, Martin A Nowak, and David G Rand. Humans display a ‘cooperative phenotype’ that is domain general and temporally stable. *Nature Communications*, 5, 2014.
- [37] D. G. Rand, A. Peysakhovich, G. T. Kraft-Todd, G. E. Newman, O. Wurzbacher, M. A. Nowak, and J. D. Greene. Social heuristics shape intuitive cooperation. *Nature Communications*, 5:3677, 2014.
- [38] David G Rand. The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology*, 299:172–179, 2012.
- [39] David G Rand and Martin A Nowak. Human cooperation. *Trends in Cognitive Sciences*, 17(8):413–425, 2013.
- [40] David G Rand, Corina E Tarnita, Hisashi Ohtsuki, and Martin A Nowak. Evolution of fairness in the one-shot anonymous ultimatum game. *Proceedings of the National Academy of Sciences*, 110(7):2581–2586, 2013.
- [41] Alvin E Roth, Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir. Bargaining and market behavior in jerusalem, ljubljana, pittsburgh, and tokyo: An experimental study. *The American Economic Review*, 81(5):1068–1095, 1991.
- [42] Siddharth Suri and Duncan J Watts. Cooperation and contagion in web-based, networked public goods experiments. *ACM SIGecom Exchanges*, 10(2):3–8, 2011.
- [43] Björn Wallace, David Cesarini, Paul Lichtenstein, and Magnus Johannesson. Heritability of ultimatum game responder behavior. *Proceedings of the National Academy of Sciences*, 104(40):15631–15634, 2007.
- [44] Jordan S Wells and David G Rand. Strategic self-interest can explain seemingly ‘fair’ offers in the ultimatum game. *Available at SSRN 2136707*, 2013.

- [45] Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics bulletin*, pages 80–83, 1945.
- [46] Paul J Zak, Angela A Stanton, and Sheila Ahmadi. Oxytocin increases generosity in humans. *PLoS ONE*, 2(11):e1128, 2007.

Appendix

Technical details

According to Fehr and Schmidt’s model, the utility for player i when player i receives x_i and player $-i$ receives x_{-i} is:

$$u_i(x_i, x_{-i}) = x_i - \alpha_i \max(x_{-i} - x_i, 0) - \beta_i \max(x_i - x_{-i}, 0), \quad (11)$$

where $0 \leq \beta_i \leq \alpha_i$ are individual parameters, representing the extent to which player i regrets to be worse off (parameter α_i) and better off (parameter β_i) than the other player. We say that a player has type (α, β) if he plays according to Fehr and Schmidt’s utility function with parameters α and β .

Proposition 6.1. *Consider the Ultimatum Game where the proposer has a budget of $40c$ and the Dictator Game where the dictator has a budget of $30c$ and the recipient already owns $10c$. The following statements hold:*

1. *If the proposer has type $(\alpha, \beta) = (0, 0)$ and believes he is playing with a recipient of type $(\alpha, \beta) = (0, 0)$, then he would donate nothing in the DG and offer nothing in the UG.*
2. *If the proposer has type $(\alpha, \beta) = (0, 0)$ and believes he is playing with a recipient of type $(\alpha, \beta) = (0.5, 0.25)$, then he would donate nothing in the DG and offer $10c$ in the UG.*
3. *If the proposer has type $(\alpha, \beta) = (0, 0)$ and believes he is playing with a recipient of type $(\alpha, \beta) = (2, 0.6)$, then he would donate nothing in the DG and offer $16c$ in the UG.*
4. *If the proposer has type $(\alpha, \beta) = (0, 0)$ and believes he is playing with a recipient of type $(\alpha, \beta) = (4, 0.6)$, then he would donate nothing in the DG and offer $17.78c$ in the UG.*

Proof. The first statement is trivial. As for the second one, the part concerning the DG is trivial. As for the part concerning the UG, we first observe that since the proposer has type $(0, 0)$ then he makes the minimum offer such that the recipient accepts. Now, observe that rejecting an offer gives utility 0 to the recipient, while accepting an offer gives utility negative only if that offer x is smaller than $10c$, in which case the recipient gets a utility of $2x - 20 < 0$. Thus, the proposer offers $10c$ and this completes the proof of the second statement. The proof of the third and fourth statements are very much alike that of the second statement. \square

Experimental instructions

Here we report full experimental instructions of Treatment 1. Instructions of Treatment 2 were exactly the same, apart from the order.

Screen 1.

This is the first part of the HIT. You have been paired with another, anonymous MTurk worker. Your bonus will depend on your and his choices.

We give you a budget of 40 cents. Decide how much, if any, to offer to the other participant. At the same time, he will decide his minimal acceptable offer.

If your offer is larger than or equal to his minimal acceptable offer, then he will get your offer and you will keep the rest. Otherwise, if your offer is smaller than his minimal acceptable offer, both of you get nothing.

The other person is real and will really make a decision.

Screen 2.

Now we will ask you some comprehension questions to make sure you understand the game. You MUST answer all questions CORRECTLY to receive the payment. In case you fail one of these questions, you will immediately see the message "We thank you for your time spent taking this survey. Your response has been recorded", and you will not receive any payment.

Screen 3.

Assume YOU offer 12 cents and the minimal acceptable offer of the OTHER PARTICIPANT is 10 cents. What is your bonus? (Here participants could choose between 0c, 10c, 12c, and 28c).

Screen 4.

Assume YOU offer 12 cents and the minimal acceptable offer of the OTHER PARTICIPANT is 10 cents. What is his bonus? (Here participants could choose between 0c, 10c, 12c, and 28c).

Screen 5.

Assume YOU offer 8 cents and the minimal acceptable offer of the OTHER PARTICIPANT is 14 cents. What is your bonus? (Here participants could choose between 0c, 8c, 14c, and 32c).

Screen 6.

Assume YOU offer 8 cents and the minimal acceptable offer of the OTHER PARTICIPANT is 14 cents. What is his bonus? (Here participants could choose between 0c, 8c, 14c, and 32c).

Screen 7.

Congratulations! You have passed the comprehension questions. Now, it's time to make your decision. WHAT IS YOUR OFFER? (Here participants could select every *even* amount of money between 0c and 40).

Screen 8.

This is the second part of the HIT. Now, you have been paired with ANOTHER anonymous MTurker. This time your bonus depends ONLY on your own choice: the other participant CANNOT influence your bonus.

You are given a new budget of 30 cents and the other participant is given 10 cents. You have to decide how much, if at all, to DONATE to the other participant. The other participant has no say and will accept any donation.

Note that the decision is unilateral: nobody will have the choice to make a donation to you.

The other person is real and will really get your donation.

Screen 9.

Now we will ask you some comprehension questions to make sure you understand the game. You MUST answer all questions CORRECTLY to receive the payment. In case you fail one of these questions, you will immediately see the message "We thank you for your time spent taking this survey. Your response has been recorded", and you will not receive any payment.

Screen 10.

If you donate 10 cents, what is YOUR bonus? (Here participants could choose between 30c, 20c, 10c, and 0c).

Screen 11.

If you donate 10 cents, what is the OTHER PARTICIPANT's bonus? (Here participants could choose between 30c, 20c, 10c, and 0c).

Screen 12.

If you donate 0 cents, what is YOUR bonus? (Here participants could choose between 30c, 20c, 10c, and 0c).

Screen 13.

If you donate 0 cents, what is THE OTHER PARTICIPANT's bonus? (Here participants could choose between 30c, 20c, 10c, and 0c).

Screen 14.

If you donate 12 cents, what is YOUR bonus? (Here participants should choose between 22c, 18c, 8c, and 12c).

Screen 15.

If you donate 12 cents, what is the OTHER PARTICIPANT's bonus? (Here participants should choose between 22c, 18c, 8c, and 12c).

Screen 16.

Congratulations! You have passed the comprehension questions. Now, it's time to make your decision.
WHAT IS YOUR DONATION?